



## RESGen: Renewable Energy Scenario Generation Platform

Iversen, Jan Emil Banning; Pinson, Pierre

*Published in:*  
Proceedings of IEEE PES General Meeting

*Publication date:*  
2016

*Document Version*  
Peer reviewed version

[Link back to DTU Orbit](#)

*Citation (APA):*  
Iversen, J. E. B., & Pinson, P. (2016). RESGen: Renewable Energy Scenario Generation Platform. In *Proceedings of IEEE PES General Meeting* IEEE.

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# RESGen: Renewable Energy Scenario Generation Platform

Emil B. Iversen, and Pierre Pinson, *Senior Member, IEEE*

**Abstract**—Space-time scenarios of renewable power generation are increasingly used as input to decision-making in operational problems. They may also be used in planning studies to account for the inherent uncertainty in operations. Similarly using scenarios to derive chance-constraints or robust optimization sets for corresponding optimization problems is useful in a power system context. Generating and evaluating such space-time scenarios is difficult. While quite a number of proposals have appeared in the literature, a gap between methodological proposals and actual usage in operational and planning studies remains. Consequently, our aim here is to propose an open-source platform for space-time probabilistic forecasting of renewable energy generation (wind and solar power). This document covers both methodological and implementation aspects, to be seen as a companion document for the open-source scenario generation platform. It can generate predictive densities, trajectories and space-time interdependencies for renewable energy generation. The underlying model works as a post-processing of point forecasts. For illustration, two setups are considered: the case of day-ahead forecasts to be issued once a day, and for rolling windows with regular updates, with application to the western part of the United States, with both wind and solar power generation.

**Index Terms**—Spatio-temporal forecasting, Probabilistic forecasting, Scenario generation, Renewable energy, Quantile regression, Copula

## I. INTRODUCTION

Energy generated from renewable sources has attracted attention in recent years as a sustainable solution to the world's growing energy demand. While hydro is easily dispatched, provided the reservoirs are full, wind and solar power are intermittent and uncertain and thus provides challenges for their successful integration into the energy system. To partly offset or mitigate these issues, accurate forecasts of future wind and solar power generation are required. For an efficient utilization of wind and solar power, in terms of economic costs as well as grid stability, forecasts providing the full predictive density of wind and solar power generation are to be preferred ([1], [2]). These types of forecasts are referred to as *probabilistic* forecasts. They stand in contrast to *point* forecasts, which are single-valued, typically informing about the expected value of power generation per location, energy type and lead time.

Energy systems are distributed in nature. Conventionally, on the demand side, power is consumed at numerous distinct locations ranging from households to factories, typically distributed over a large geographical area. Renewable energy generation is also distributed, ranging from rooftop photovoltaic installations to wind turbines spread out across the countryside or at sea. Further, renewable generation in adjacent locations may be highly correlated as it is governed by the same weather system [3]. Thus, in view of the distributed nature of supply

and demand, combined with network constraints, probabilistic spatio-temporal forecasts comprise optimal input to decision-making problems.

While probabilistic spatio-temporal forecasting of renewable energy generation is novel, quite a bit has been done already on the basics of wind power forecasting. The propagation of wind power forecast errors has been studied in [10] and [11], where it is suggested that a spatio-temporal model would allow producing better forecasts. In [12] an ellipsoidal uncertainty set is used to capture correlation between wind farms at different locations. Very-short-term forecasting of wind power generation has been considered in [13] using a sparse vector auto regressive model to forecast the power production at 22 different sites. A vector auto regressive model is also considered in [14], however, the focus is on providing better point forecasts. Copula models were employed in [15], forecasting densities of weather variables, in [16], where a machine learning and copula based approach is used to predict wind power, and in [17] where a copula approach is used to model wind power prediction errors while reducing the parameters needed through a precision matrix formulation. The relevant challenges with probabilistic renewable energy forecasting include (i) the dynamic and time-varying nature of the weather, and (ii) the nonlinear and bounded nature of the power conversion process. Spatio-temporal probabilistic forecasting is further complicated by the large data sets and the dimension of the models, which in turn can lead to large parameter spaces and high computational costs.

In this work we describe a generic tool for obtaining probabilistic spatio-temporal forecasts for renewable energy generation, within a modular framework. The tool makes it possible to extract predictive quantiles, moments, prediction intervals, predictive densities and to simulate trajectories. The prediction tool is coded in python as many non-statisticians are more familiar with this language as opposed statistical software such as R or SAS. This is intended to allow for people with high programming skills to tweak and improve upon the code without having an in-depth knowledge of statistics. The code provided is open source and free to used under an open source BSD 3-clause. The remainder of the document is structured as follows. In section 2 the data used for this study is detailed. Next, section 3 gives a brief introduction to the mathematics underlying the prediction tool. Section 4 highlights how to run the code in python. In section 5 the outputs of the model are considered. Finally section 6 concludes the paper.

## II. DATA

The data used in this study covers wind and solar power generation for the Western Interconnection in the United

States, which is shown in Figure 1.

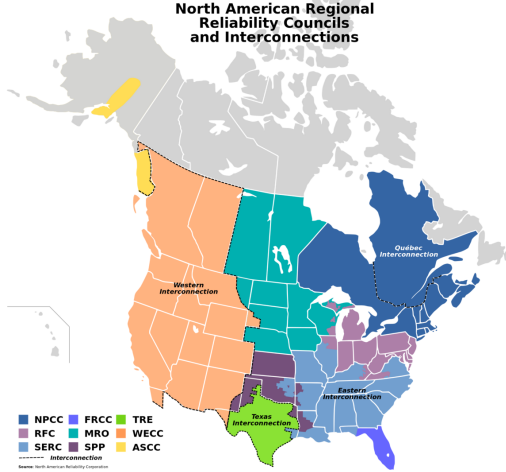


Fig. 1. Interconnections, United States

Within the Western Interconnection a total of 39 load regions exist. In Figure 2 the approximate location of these different load regions are shown [18]. Only a subset of the load regions are present in the wind or solar data sets.

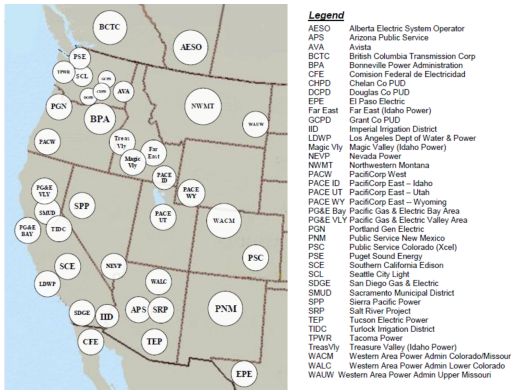


Fig. 2. Load regions, Western Interconnection

The data contains aggregate power generation from wind and solar. There are 22 load regions for wind and 26 load regions for solar present in the data. Further the data also contains various forecasts for power generation. These include day-ahead forecasts with an hourly time resolution spanning the next day and intra-day forecast spanning 15-min ahead to 5-hour ahead with a 15 minute time resolution. Each of the data sets contain one year of data.

The data sets used in this study were of high quality. While the quality of the data is high, the same conclusion cannot be reached for the quality of the point forecasts used in this study. This may to some degree be explain by the forecast being computed before a market gate closure. It is clear that the point forecasts do not make use of past observations of generated power to predict the following day. Thus, there is a clear avenue for improvements to yield a better predictive performance.

### III. MATHEMATICAL FOUNDATIONS

The forecasting tool provides a non-parametric approach to spatio-temporal probabilistic forecasting. Quantile regression

is used to estimate marginal predictive densities and a copula is subsequently applied to model the interdependence in time and space. The implementation is not dependent on specific data sets, time resolution, locations or forecast horizons and the approach can be readily adapted to other data sets. In this section a brief overview of the mathematics behind the implementation is presented.

#### A. Quantile regression density estimation

Suppose  $X$  is a real valued random variable with cumulative distribution function  $F_X(x) = P(X \leq x)$ . The  $\tau$ 'th quantile of  $X$  is given as

$$Q_X(\tau) = F_X^{-1}(\tau) = \inf \{x : F_X(x) \geq \tau\} \quad (1)$$

where  $\tau \in [0, 1]$ .

Next define the loss function as  $\rho_\tau(x) = |x(\tau - \mathbb{I}_{(x < 0)})|$ , where  $\mathbb{I}$  is an indicator variable, that is one when the condition is met and zero otherwise. A distinct quantile can be obtained by minimizing the expected loss of  $X - u$  with regards to  $u$ :

$$\begin{aligned} & \min_u E(\rho_\tau(X - u)) \\ &= \min_u (\tau - 1) \int_{-\infty}^u (x - u) dF_X(x) \\ & \quad + \tau \int_u^{\infty} (x - u) dF_X(x). \end{aligned} \quad (2)$$

Set the derivative of the expected loss function to 0 and let  $q_\tau$  be the solution of the following equation,

$$0 = (1 - \tau) \int_{-\infty}^{q_\tau} dF_X(x) - \tau \int_{q_\tau}^{\infty} dF_X(x). \quad (3)$$

This equation reduces to

$$F_X(q_\tau) = \tau. \quad (4)$$

Thus  $q_\tau$  is  $\tau$ 'th quantile of the stochastic variable  $X$ .

The  $\tau$  sample quantile can be found by solving the minimization problem:

$$\begin{aligned} \hat{q}_\tau &= \arg \min_{q \in R} \sum_{i=1}^n \rho_\tau(x_i - q) \\ &= \arg \min_{q \in R} \left[ (1 - \tau) \sum_{x_i < q} (x_i - q) + \tau \sum_{x_i \geq q} (x_i - q) \right]. \end{aligned} \quad (5)$$

Solving this minimization problem for a suitable number of quantiles the marginal cumulative density function and its inverse can now be estimated. Notice here, however, that this formulation does not allow for the density to depend on external regressors. This is remedied by introducing the conditional quantile regression.

#### B. Conditional quantile regression density estimation

Let the  $\tau$ 'th conditional quantile function be given by  $Q_{X|Z}(\tau) = Z\beta_\tau$ . Given the distribution function of  $X$ ,  $\beta_\tau$  can be obtained by solving

$$\beta_\tau = \arg \min_{\beta \in \mathbb{R}^k} E(\rho_\tau(X - Z\beta)). \quad (6)$$

The sample analog can be solved, which yields the estimator of  $\beta_\tau$ , i.e.,

$$\hat{\beta}_\tau = \arg \min_{\beta \in \mathbb{R}^k} \sum_{i=1}^n (\rho_\tau(X_i - Z\beta)). \quad (7)$$

Solving this minimization problem for a suitable number of quantiles we obtain estimates of parameter values in the model for each specific quantile. Again, collecting these quantiles the marginal cumulative density functions can be estimated. While in this specific case  $Q_{X|Z}(\tau)$  is linear this is in general not a requirement.

### C. Copula estimation

Consider a random vector  $(X_1, X_2, \dots, X_d)$ . Suppose its margins are continuous, i.e. the marginal CDFs  $F_i(x) = \mathbb{P}[X_i \leq x]$  are continuous functions. By applying the probability integral transform to each component, the random vector

$$(U_1, U_2, \dots, U_d) = (F_1(X_1), F_2(X_2), \dots, F_d(X_d)) \quad (8)$$

has uniformly distributed marginals.

The copula of  $(X_1, X_2, \dots, X_d)$  is defined as the joint cumulative distribution function of  $(U_1, U_2, \dots, U_d)$ :

$$C(u_1, u_2, \dots, u_d) = \mathbb{P}[U_1 \leq u_1, U_2 \leq u_2, \dots, U_d \leq u_d]. \quad (9)$$

The copula  $C$  contains all information on the dependence structure between the components of  $(X_1, X_2, \dots, X_d)$  whereas the marginal cumulative distribution functions  $F_i$  contain all information on the marginal distributions.

The importance of the above is that the reverse of these steps can be used to generate pseudo-random samples from general classes of multivariate probability distributions. That is, given a procedure to generate a sample  $(U_1, U_2, \dots, U_d)$  from the copula distribution, the required sample can be constructed as

$$(X_1, X_2, \dots, X_d) = (F_1^{-1}(U_1), F_2^{-1}(U_2), \dots, F_d^{-1}(U_d)). \quad (10)$$

The inverses  $F_i^{-1}$  are unproblematic as the  $F_i$  were assumed to be continuous. The above formula for the copula function can be rewritten to correspond to this as:

$$C(u_1, u_2, \dots, u_d) = \mathbb{P}[X_1 \leq F_1^{-1}(u_1), X_2 \leq F_2^{-1}(u_2), \dots, X_d \leq F_d^{-1}(u_d)]. \quad (11)$$

In this implementation a Gaussian copula is used. This copula is particularly simple to estimate due to the fact that it is completely characterized by the covariance matrix of the associated multivariate normal distribution. We have that

$$(\Phi^{-1}(U_1), \Phi^{-1}(U_2), \dots, \Phi^{-1}(U_d)) \sim \mathcal{N}_{\mathbb{R}^d}(0, \Sigma), \quad (12)$$

where  $\Phi^{-1}$  is the inverse cumulative density function for the standard normal distribution and  $\Sigma$  describes the covariance matrix that characterizes the Gaussian copula. Thus we can estimate  $\Sigma$  with a method for estimating the covariance matrix for a normal distribution. The sample covariance matrix is given by

$$\hat{\Sigma} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T, \quad (13)$$

where  $x_i$  is the  $i$ -th observation of the  $d$ -dimensional random vector, and

$$\bar{x} = \begin{bmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_d \end{bmatrix} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (14)$$

is the sample mean.

## IV. PYTHON IMPLEMENTATION

The model implementation consists of two models, one day-ahead model, predicting hourly power production for the coming day, and one intra-day model, predicting 15 minute power production up to the following 5 hours. The main difference between the two models are that the day-ahead model constructs a marginal density for every hour of the coming day whereas the intra-day model de-seasons the data and constructs the marginal densities for each lead time instead. This section explains how to run the different models, however further details of the code functionality is included in the code as comments and the authors suggest looking through these comments for further details of the actual functionality.

### A. The Sample Data

The day-ahead model (either wind or solar power) uses hourly observation of generated power along with day-ahead forecasts of power generation. The forecasts consist of one point forecasts for every location for every hour of the day.

The intra-day model (either wind or solar power) uses 15-minute observations of generated power along with intra-day forecasts of power generation. The forecasts for the intra-day model overlap such that every 15 minutes point forecasts are generated for the next 5 hours with a time resolution of 15 minutes.

### B. The Day-Ahead Model

A short introduction to running the day-ahead model is provided. Loading the data is done by the python script `ReadingData.py`. The model is fitted to the data by the python script `ConditionalQuantileRegression.py`. The predictive quantiles for the generated power are estimated and are then used to construct a predictive cumulative density function and its inverse. To obtain uniform marginals, the generated power observations are transformed by the predictive cumulative density functions. These uniform marginals are then converted to standard normal marginals and the covariance matrix is estimated as the sample covariance. Thus the model is specified.

In Figure 3 the estimated correlation between different locations at different hours of the day are shown for the day-ahead solar model. The correlation matrix is ordered such that location 1 hours 1-24 is in the top left, below that is location 2 hours 1-24 and so on.

The day ahead model can be simulated by the python script `SimulationConditionalQuantileRegression.py`. This is done by simulating normally distributed variables with the estimated interdependence structure and reversing the procedure specified above.

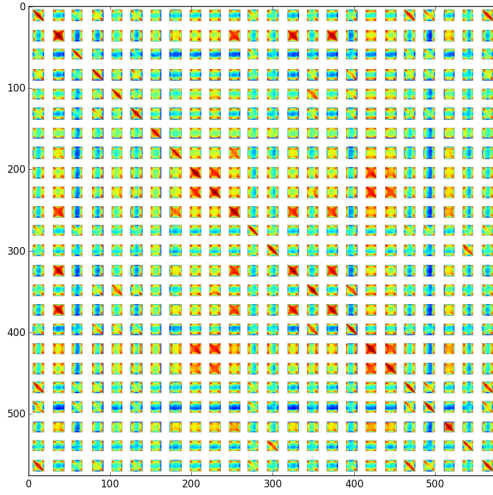


Fig. 3. Correlation matrix describing the interdependence structure. Warmer colors equal a high correlation.

### C. The Intra-Day Rolling Model

A short introduction to running the intraday model is provided. The data is loaded by the python script `ReadingDataRolling.py`. The model is fitted to the data by the python script `RollingQuantileRegression.py`. First the observations and predictions are transformed to a stationary domain. The predictive quantiles for the transformed observations are estimated via quantile regression which in turn are used to construct a predictive cumulative density function and its inverse. To obtain uniform marginals, the transformed observation on the stationary domain are transformed once more by the predictive cumulative density functions. These uniform marginals are then converted to standard normal marginals and the covariance matrix is estimated as the sample covariance. Thus the model is specified.

The day ahead model can be simulated by the python script `SimulationRollingQuantileRegression.py`. First the uniforms with the proper interdependence structure is simulated. These are then converted to de-seasoned domain by using the inverse predictive cumulative density functions estimated via quantile regression. Further the predictions on the de-seasoned domain are transformed to the original power domain via a transformation. The intra-day model is different from the day-ahead model in the sense that the forecasts are overlapping. The simulated trajectories are further consider in the model output section.

## V. MODEL OUTPUTS

Model outputs that can be extracted from the models include quantiles, moments, prediction intervals, predictive densities and simulated trajectories. Here we focus on the simulated trajectories.

### A. Day-Ahead Trajectories

The day-ahead simulated trajectories span the next day with an hourly time resolution. In Figure 4 such trajectories are shown for wind power generation at the location APS. Along with the 3 simulated trajectories in blue the predicted power is shown in green. The actual realization is shown in red. Notice how both the simulations and the realization deviate quite a

lot from the prediction already in the initial hour. This is most likely due to the quality of the prediction, in that it does not use past observations of power generation to predict the output power. As such the models could be substantially improved by a better point prediction of the generated wind power. Notice also the auto-correlation in the residuals from the predicted power which is an important feature for simulation tool as to provide reliable multi-horizon trajectories.

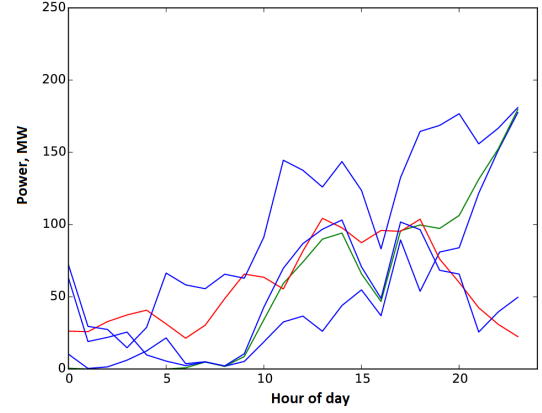


Fig. 4. Day-ahead wind power generation is shown on vertical axis and hour on horizontal axis. Prediction in green, scenarios in blue and realization in red.

Similarly for solar power, Figure 5 represent simulated trajectories for day-ahead solar power generation for the location APS. Again the simulations are in blue, the prediction is in green and the realization is in red. Notice here how the diurnal effect is captured by the simulated trajectories. The yearly seasonality enters indirectly into the model through the point predictions of solar power generation.

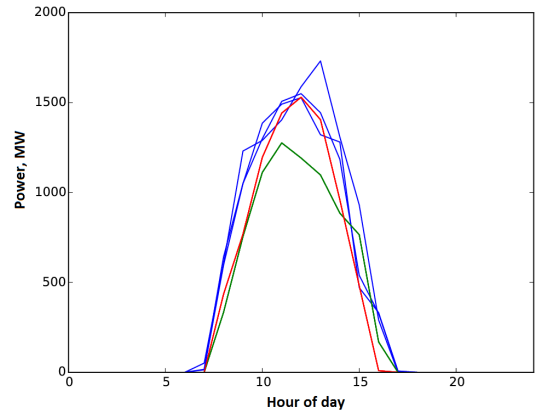


Fig. 5. Day-ahead solar power generation is shown on vertical axis and hour on horizontal axis. Prediction in green, scenarios in blue and realization in red.

### B. Intra-Day Trajectories

The intra-day simulated trajectories span the following 15 minutes to 5 hours with a time resolution of 15 minutes. In Figure 6 simulate trajectories are shown for wind power generation at the location APS. Three simulated trajectories are shown in blue, the predicted power in green and the actual realization is shown in red. As for the day-ahead models, the simulations and the realization deviate quite a lot from the prediction already in the first step, that is 15 minutes ahead. Again the likely explanation is the quality of the point prediction, in that it does not use past observations of power generation to predict the output power nor is it post-processed

in a statistical sense to correct for possible bias. As for the day-ahead models, the intra-day models predictive performance could be substantially improved by better point predictions of generated power. Notice again the auto-correlation in the residuals from the predicted power.

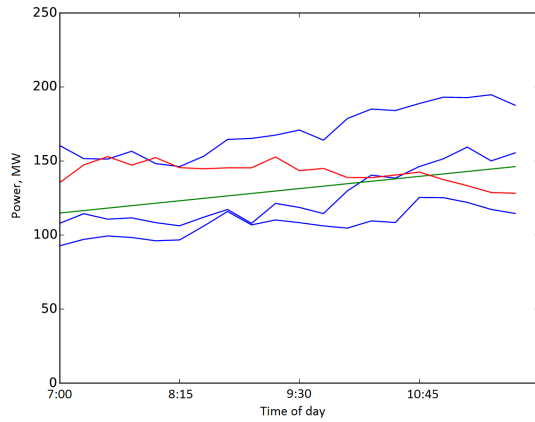


Fig. 6. Intraday wind power generation is shown on vertical axis and lead time on horizontal axis. Prediction in green, scenarios in blue and realization in red.

Simulations from the intra-day solar power model are shown in Figure 7 for the location APS. Again the simulations are in blue, the prediction is in green and the realization is in red. In Figure 7 the trajectories simulation starts at 7:00 am and spans the next 5 hours. We see that the diurnal variation is captured nicely by the model.

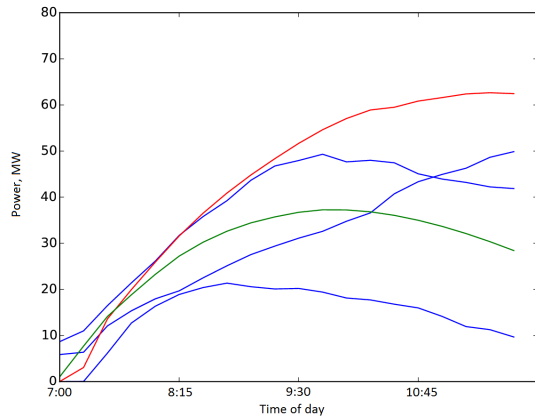


Fig. 7. Intraday solar power generation is shown on vertical axis and lead time on horizontal axis. Prediction in green, scenarios in blue and realization in red.

## VI. CONCLUSION

The work presented here provides a generic, accessible and modular framework for providing probabilistic spatio-temporal forecasts. From this model it is possible to extract predictive quantiles, moments, prediction intervals, predictive densities and to simulate trajectories, all in a setting that preserves the distributions of and spatio-temporal dependence in the observations. A versatile and state-of-the-art tool for probabilistic spatio-temporal forecasting is provided. The tool provides models for day-ahead as well as intra-day forecasts for both wind and solar power. The model is provided along with a data set pertaining to the Western Interconnection in the United States. However, the implementation is not dependent on specific data sets, time resolution, locations or forecast

horizons and the approach can be readily adapted to other data sets. The code provided is open source and free to use under an open source BSD 3-clause.

## REFERENCES

- [1] J. M. Morales, A. J. Conejo, H. Madsen, P. Pinson, and M. Zugno, *Integrating Renewables in Electricity Markets – Operational Problems*, ser. International Series in Operations Research & Management Science. New York: Springer, 2014, vol. 205.
- [2] P. Pinson, “Wind energy: Forecasting challenges for its operational management,” *Statistical Science*, vol. 28, no. 4, pp. 564–585, 2013.
- [3] T. Ackermann, *Wind power in power systems*. John Wiley & Sons, 2005.
- [4] V. Chew, “Simultaneous prediction intervals,” *Technometrics*, vol. 10, no. 2, pp. 323–330, 1968.
- [5] N. Ravishanker, L. Wu, and J. Glaz, “Multiple prediction intervals for time series: comparison of simultaneous and marginal intervals,” *Journal of Forecasting*, vol. 10, no. 5, pp. 445–463, 1991.
- [6] Ö. Jordá and M. Marcellino, “Path forecast evaluation,” *Journal of Applied Econometrics*, vol. 25, no. 4, pp. 635–662, 2010.
- [7] Ö. Jordá, M. Knüppel, and M. Marcellino, “Empirical simultaneous prediction regions for path-forecasts,” *International Journal of Forecasting*, vol. 29, no. 3, pp. 456–468, 2013.
- [8] J. Jeon and J. W. Taylor, “Using conditional kernel density estimation for wind power density forecasting,” *Journal of the American Statistical Association*, vol. 107, no. 497, pp. 66–79, 2012.
- [9] P. Li, X. Wang, and X. Wang, “Review on probabilistic forecasting of wind power generation,” *Renewable and Sustainable Energy Reviews*, vol. 32, pp. 255–270, 2014.
- [10] R. Girard and D. Allard, “Spatio-temporal propagation of wind power prediction errors,” *Wind Energy*, vol. 16, no. 7, pp. 999–1012, 2013.
- [11] J. Tastu, P. Pinson, E. Kotwa, H. Madsen, and H. A. Nielsen, “Spatio-temporal analysis and modeling of short-term wind power forecast errors,” *Wind Energy*, vol. 14, no. 1, pp. 43–60, 2011.
- [12] P. Li, X. Guan, J. Wu, and X. Zhou, “Modeling dynamic spatial correlations of geographically distributed wind farms and constructing ellipsoidal uncertainty sets for optimization-based generation scheduling,” 2015.
- [13] J. Dowell and P. Pinson, “Very-short-term probabilistic wind power forecasts by sparse vector autoregression,” 2015.
- [14] R. J. Bessa, A. Trindade, and V. Miranda, “Spatial-temporal solar power forecasting for smart grids,” *Industrial Informatics, IEEE Transactions on*, vol. 11, no. 1, pp. 232–241, 2015.
- [15] R. Schefzik, T. L. Thorarinsdottir, T. Gneiting *et al.*, “Uncertainty quantification in complex simulation models using ensemble copula coupling,” *Statistical Science*, vol. 28, no. 4, pp. 616–640, 2013.
- [16] M. Wytock and J. Z. Kolter, “Large-scale probabilistic forecasting in energy systems using sparse gaussian conditional random fields,” in *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*. IEEE, 2013, pp. 1019–1024.
- [17] J. Tastu, P. Pinson, and H. Madsen, “Space-time trajectories of wind power generation: Parameterized precision matrices under a gaussian copula approach,” *Lecture Notes in Statistics: Modeling and Stochastic Learning for Forecasting in High Dimension*, 2014.
- [18] T. Guo, G. Liu, and J. Toolson, “Balancing authority cooperation concepts to reduce variable generation integration costs in the western interconnection: Intra-hour scheduling,” 2013.
- [19] E. B. Iversen, J. M. Morales, J. K. Møller, and H. Madsen, “Probabilistic Forecasts of Solar Irradiance by Stochastic Differential Equations,” *Environmetrics*, vol. 25, no. 1, pp. 152–164, 2014.
- [20] U. Cherubini, E. Luciano, and W. Vecchiato, *Copula methods in finance*. John Wiley & Sons, 2004.
- [21] A. M. Foley, P. G. Leahy, A. Marvuglia, and E. J. McKeogh, “Current methods and advances in forecasting of wind power generation,” *Renewable Energy*, vol. 37, no. 1, pp. 1–8, 2012.
- [22] A. Costa, A. Crespo, J. Navarro, G. Lizcano, H. Madsen, and E. Feitosa, “A review on the young history of the wind power short-term prediction,” *Renewable and Sustainable Energy Reviews*, vol. 12, no. 6, pp. 1725–1744, 2008.
- [23] E. W. Law, A. A. Prasad, M. Kay, and R. A. Taylor, “Direct normal irradiance forecasting and its application to concentrated solar thermal output forecasting—a review,” *Solar Energy*, vol. 108, pp. 287–307, 2014.
- [24] R. H. Inman, H. T. Pedro, and C. F. Coimbra, “Solar forecasting methods for renewable energy integration,” *Progress in energy and combustion science*, vol. 39, no. 6, pp. 535–576, 2013.